MPEG 영상부호화 기법

2001. 11. 24

권 순 각(동의대학교)

1

International standards

International standards	Bit rate	sources	application
JPEG (still image)	**10 – 20:1**	**Image**	**Storage, Compressed TX**
H.261 (slow motion)	**p x 64 Kbps**	**Video**	**Videophone, Videoconference**
MPEG-1 (moving picture)	**1 – 1.5 Mbps**	**Video/audio**	**Storage (DSM)**
MPEG-2 (H.262)	**≤ 100 Mbps**	**Video/audio**	**Transmission (DTV, HDTV)**
MPEG-4 (H.263)	**≤ 4 Mbps (≤ 64 Kbps)**	**Video/audio**	**Videophone, Mobile, Internet**

2

MPEG video

- MPEG = Moving Picture Experts Group
- Part of ISO/IEC
- Aim was to create the best video compression standards for multimedia and broadcast applications
- MPEG-1 Video aimed at SIF resolution
  - 352x240, 30Hz, non-interlaced, 1.5 Mbps
  - CD-ROM applications
- MPEG-2 Video aimed at ITU-R 601 resolution
  - 720x480, 30Hz, interlaced, 4-10 Mbps
  - Enlarge applications including HDTV at less than 100Mbps
- MPEG-4 Video aimed at a number of resolutions
  - SIF, CIF etc., and arbitrary shaped objects
  - Wired(Internet) or wireless multimedia interactive applications

3

Comparison of JPEG, H.261, MPEG-1

		JPEG (Baseline)	H.261	MPEG-1
(8x8) DCT		Yes	Yes	Yes
Zigzag scan		Yes	Yes	Yes
2D runlength coding		Yes	Yes	Yes
Quantizer		Default tables	32 step sizes	Intra: similar to JPEG Inter: similar to H.261
Motion estimation		No	Forward	Forward / Backward
MV accuracy		No	Integer pel	Half pel
Rate buffer control		No	Yes	Yes
VLC	TCOEFF	Downloadable	Fixed	Superset of H.261
	MBA, CBP	No	Fixed	Same as H.261
	MV, MTYPE	No	Fixed	Superset of H.261

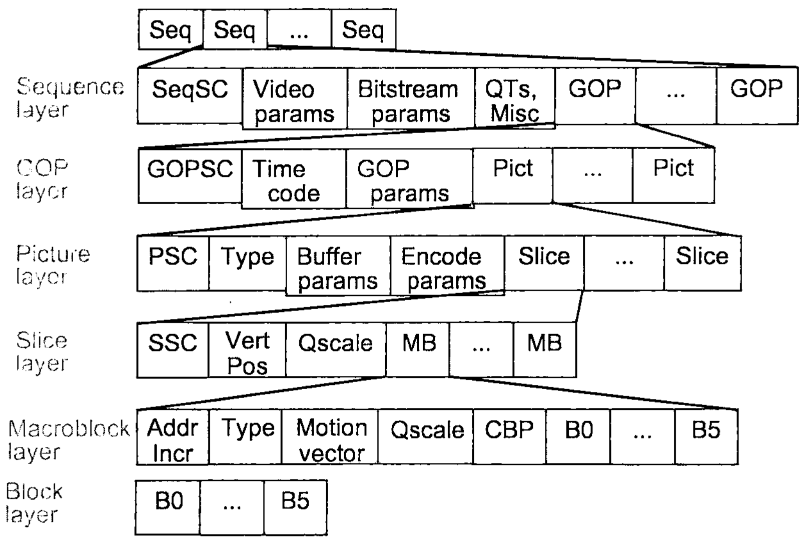
4

Video constraints for MPEG-1

constraints	parameters
Horizontal picture size	£ 768 pels
Vertical picture size	£ 576 lines
Picture area	£ 396 macroblocks
Pixel rate	£ 396 x 25 macroblocks/sec
Picture rate	£ 30 Hz
Motion vector range	£ ± 64 pels (using half-pel vectors)
Input buffer size(VBV model)	£ 327,680 bits
Bit rate	£ 1,856,000 bits/sec

5

MPEG-1 video stream



6

Functional comparison of 6 layers

- Sequence layer:
  - One or more groups of pictures
- Group of pictures layer:
  - Random access into the sequence
- Picture layer:
  - Primary coding unit
- Slice layer:
  - Resynchronization unit
- Macroblock layer:
  - Motion compensation unit
- Block layer:
  - DCT unit

7

MPEG-1 video stream

- Sequence layer
  - video params: width, height, aspect ratio, picture rate
  - Bitstream Params: bit-rate, buffer size, flag
  - QTs: for intra and inter block coding
- GOP layer
  - Time code: hour, min, sec
  - GOP params: structure of GOP
- Picture layer
  - Type: I, P, B frame
  - Buffer params: buffer status
  - Encode params: half-pel ME or not

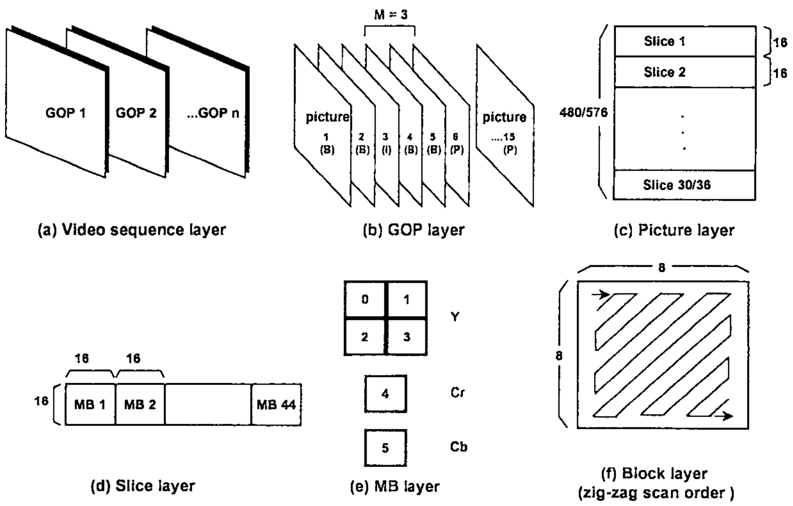
8

MPEG-1 video stream

- Slice layer
  - Vert pos: vertical position of slice
  - Qscale: quantizer scale factor
- Macroblock layer
  - Addr Incr: number of MBs to skip
  - Type: motion vector type
  - Qscale: quantizer scale factor
  - CBP: coded block pattern
- Block layer
  - DC, AC transform coefficients

9

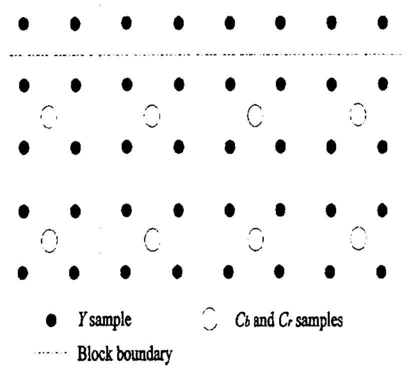
MPEG-1 data structure



10

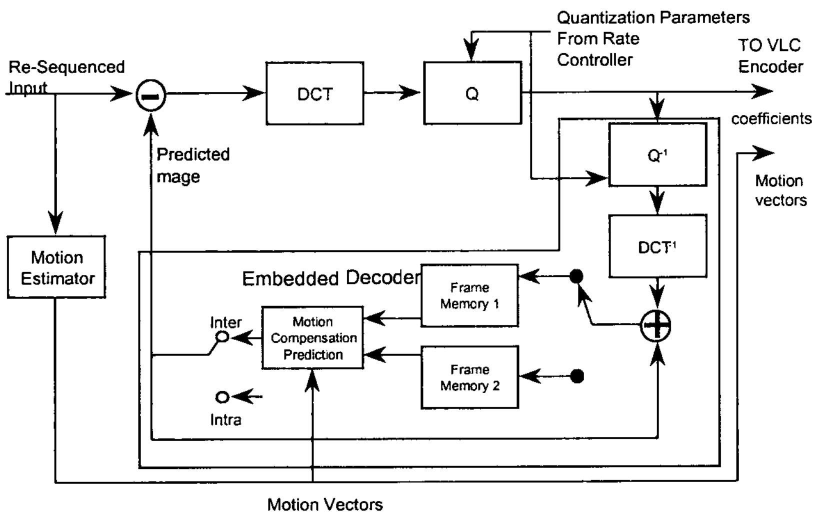
Samples of Luma & Chroma

- Reduction of spectral redundancy : downsampling
- Data sampling structure: 4:2:0 Format



11

MPEG-1 encoder



12

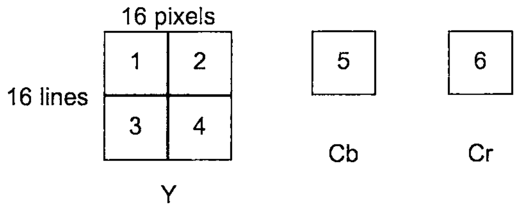
Transform

- DCT : Discrete Cosine Transform
- Reduction of spatial redundancy
- Transform block size: 8 x 8

$$f(x,y) = \frac{1}{4}\sum_{u=0}^{7}\sum_{v=0}^{7} c_u c_v f(u,v)\cos[\frac{\pi(2x+1)u}{16}]\cos[\frac{\pi(2y+1)v}{16}]$$

$$where \ u,v,x,y = 0,1,\ldots,7 \quad c_k = \begin{cases} 1/\sqrt{2}, & k = 0 \\ 1, & k \neq 0 \end{cases}$$

- MB transform order



13

DCT

- Forward N-point DCT

$$X(k) = \frac{2}{N} c_k \sum_{n=0}^{N-1} x(n)\cos\left[\frac{(2n+1)k\pi}{2N}\right]$$

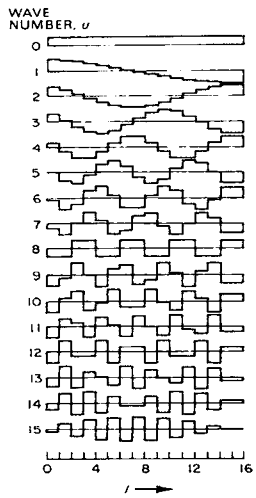
$$k = 0,1,\cdots N-1$$

- Inverse N-point DCT

$$x(n) = \sum_{k=0}^{N-1} c_k X(k)\cos\left[\frac{(2n+1)k\pi}{2N}\right]$$

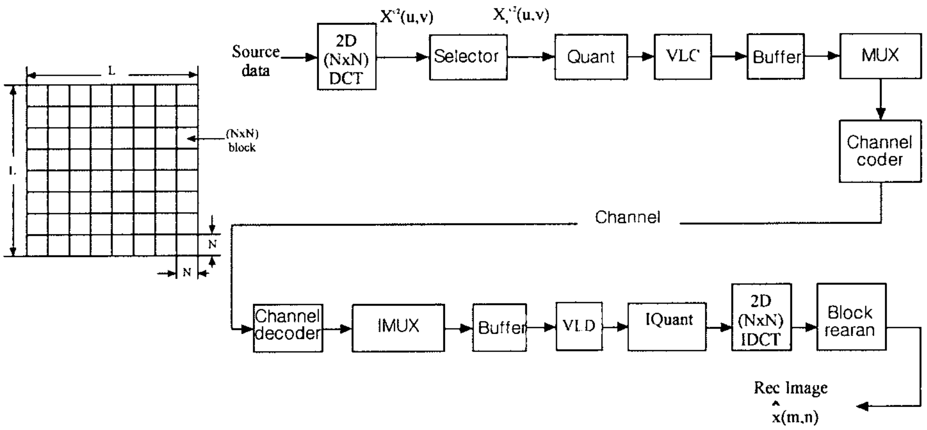
$$n = 0,1,\cdots N-1$$

$$c_k = \begin{cases} 1/\sqrt{2}, & k = 0 \\ 1, & k \neq 0 \end{cases}$$



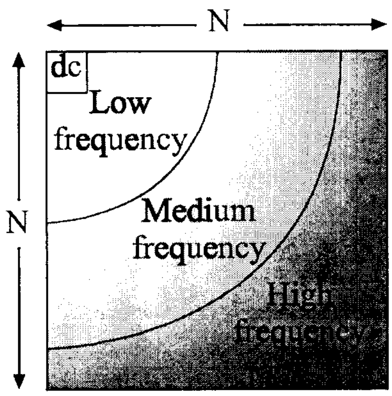
14

Block transform

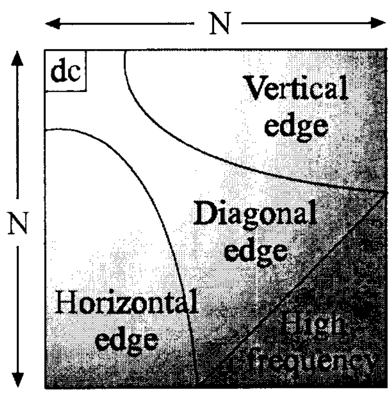


15

DCT domain



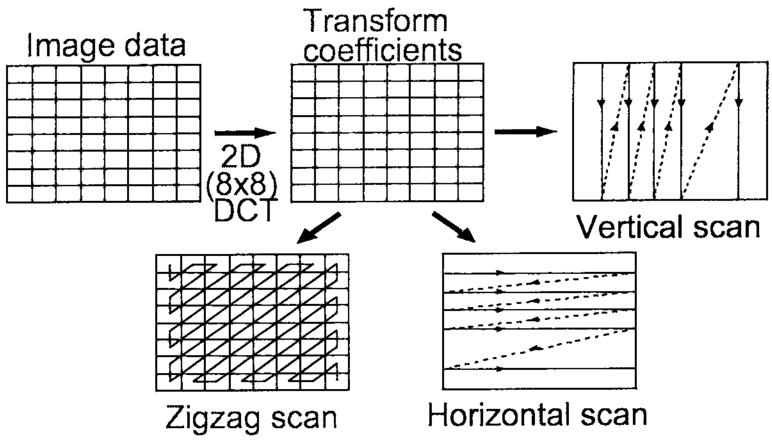
Frequency distribution



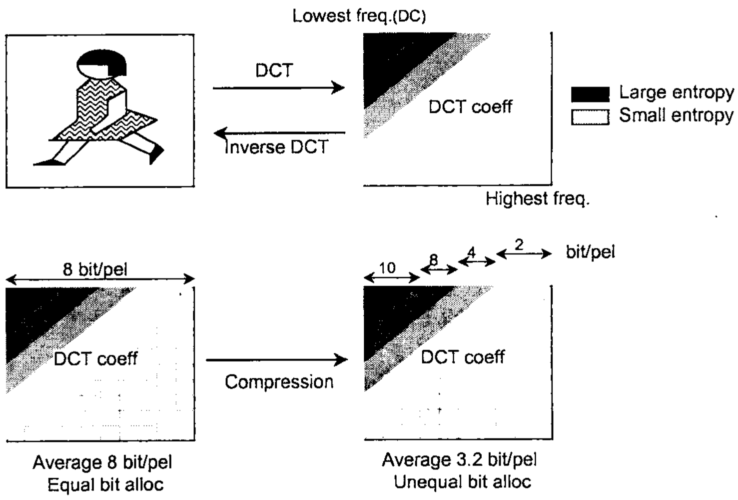
Structural decomposition

16

Scanning coefficients



Energy in DCT domain



Quantization matrix in JPEG

16	11	10	16	24	40	51	61
12	12	14	19	26	58	60	55
14	13	16	24	40	57	69	56
14	17	22	29	51	87	80	62
18	22	37	56	68	109	103	77
24	35	55	64	81	104	113	92
49	64	78	87	103	121	120	101
72	92	95	98	112	100	103	99

For luminance Y

17	18	24	47	99	99	99	99
18	21	26	66	99	99	99	99
24	26	56	99	99	99	99	99
47	66	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99
99	99	99	99	99	99	99	99

For chrominance Cb, Cr

19

Quantization matrix in MPEG

8	16	19	22	26	27	29	34
16	16	22	24	27	29	34	37
19	22	26	27	29	34	34	38
22	22	26	27	29	34	37	40
22	26	27	29	32	35	40	48
26	27	29	32	35	40	48	58
26	27	29	34	38	46	56	69
27	29	35	38	46	56	69	83

For Intra coding

16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16
16	16	16	16	16	16	16	16

For Non-intra coding

20

Quantization in JPEG/MPEG

- JPEG
  - Uniform quantizer by matrix

$$S_{quv} = \text{Nearest integer}\left(\frac{S_{uv}}{Q_{uv}}\right)$$

Inverse quantizer

$$R_{uv} = S_{quv} Q_{uv}$$

- MPEG

$$S_{quv} = 8 \times S_{uv} // (qs \times Q_{uv})$$

qs: quantizer scale range 1 to 31

//: division with rounding to nearest integer

21

Quantization (Intra)

Squv = 8 x Suv // (qs x Quv)

Ex) Let Suv = 56, qs = 2, Quv = 19,

quantized coef Squv = (8 x 56)//(2 x 19) = 12

reconstruction value is, in decoder

S'uv = (2 x Squv x qs x Quv)/16

If (S'uv & 1) = 0, S'uv = S'uv – sign(S'uv)

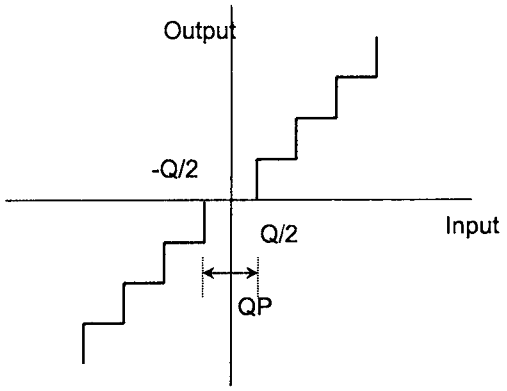
thus, S'uv = (2 x 12 x 2 x 19)/16 = 57

error is 57 – 56 =1.

22

Quantization

- INTRA AC
  - Uniform quantizer with no deadzone



23

Quantization (Non-Intra)

Squv = 8 x Suv // (qs x Quv)

Ex) Let Suv = 56, qs = 2, Quv = 19,

quantized coef Squv = (8 x 56)//(2 x 19) = 12

reconstruction value is, in decoder

S'uv = (((2 x Squv + sign(Squv))x qs x Quv)/16

If (S'uv & 1) = 0, S'uv = S'uv – sign(S'uv)

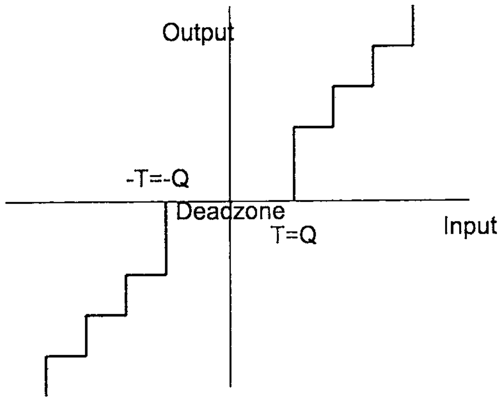
thus, S'uv = ((2 x 12 +1)x 2 x 19)/16 = 59

error is 59 – 56 =3.

24

Quantization

- INTER AC
  - Nearly uniform quantizer with deadzone
  - Input between +-T(threshold) is quantized to zero



25

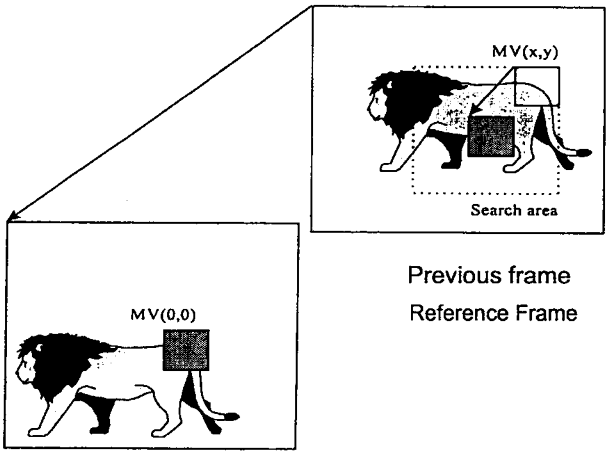
Motion Estimation & Motion Compensation

- ME: Motion Vector per luminance macroblock unit
- MC: compensate differential MB using MV
- Temporal redundancy reduction(INTER mode)
- Motion Compensated Prediction
  - P-picture: Forward Prediction
  - B-picture: Forward/Backward/Bidirectional Prediction
  - transmit difference between current frame and motion compensated previous frame

$$MB_E(t+1) = MB_F(t+1) - MB_F(t, x+dx, y+dy)$$

26

Motion Compensated Prediction



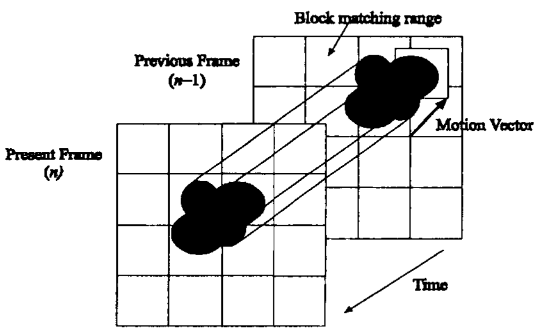
Previous frame
Reference Frame

Current frame

27

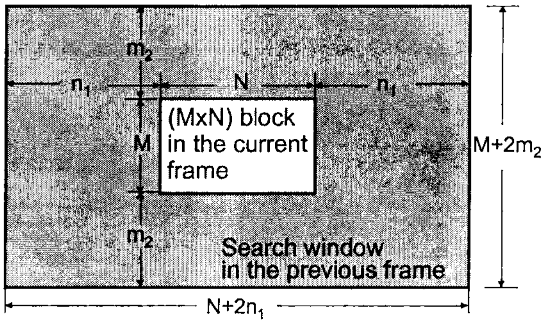
Motion estimation

- block-by-block : block matching alg.(BMA)
  full search, best match(least distortion)



28

Search range (BMA)



MV range: $\pm$ n_1 hor,
$\pm$ m_2 ver.

Computing complexity
$=(2m_2+1)\times(2n_1+1)$

29

Distortion measure

- Mean square error(MSE)

$$D(i,j)=\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}(X_{m,n}-X^{R}{}_{m+i,n+j})^2$$

$$|i|\le m_2,\qquad |j|\le n_1 \text{ in search range}$$

- Mean absolute error(MAE)

$$D(i,j)=\frac{1}{MN}\sum_{m=1}^{M}\sum_{n=1}^{N}\left|X_{m,n}-X^{R}{}_{m+i,n+j}\right|$$

This is simple, preferred.

30

example

Given search window 4x4,
Derive a motin vector for 2x2 center block.

Current block : $\begin{bmatrix} 7 & 3 & 4 & 9 \\ 5 & 8 & 6 & 4 \\ 8 & 5 & 8 & 6 \\ 6 & 9 & 5 & 2 \end{bmatrix}$

Previous block : $\begin{bmatrix} 5 & 4 & 7 & 5 \\ 7 & 9 & 5 & 9 \\ 3 & 3 & 8 & 6 \\ 8 & 6 & 7 & 3 \end{bmatrix}$

SAD result

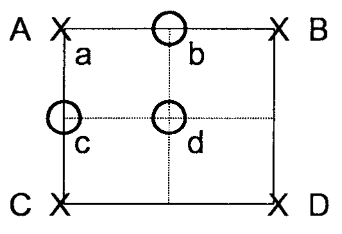
$\begin{bmatrix} 8 & 12 & 3 \\ 11 & 4 & 11 \\ 13 & 9 & 7 \end{bmatrix}$

MV = (+1, −1)

31

Half-pel ME

- Bilinear interpolation



a=A
b=(A+B)/2
c=(A+C)/2
d=(A+B+C+D)/4

X: integer pel position
O: half-pel position

- Half-pel ME

Interpola-tion → Integer-pel ME → Decima-tion

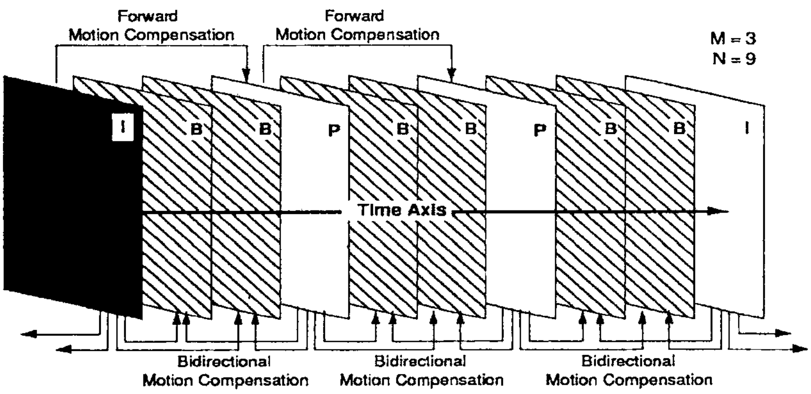
32

Forward/Backward prediction

- I: intra picture/frame
  - first frame in a GOP, coded by DCT
- P: predicted forward
  - reference is a previous frame
- B: predicted bidirectionally
  - two references (previous and future)

33

Forward/Backward prediction



GOP structure ex.: IBBPBBPBB (M=3, N=9)

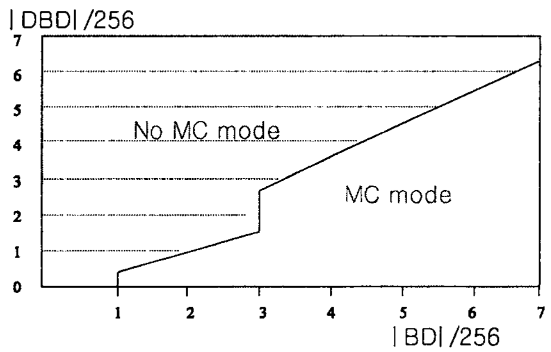
34

Selection of macroblock type

- MC or No MC
  - If motion compensation is best, select "MC" and transmit motion vector
  - For B-picture, select forward, backward or interpolated
  - Otherwise, select "No MC"; do not transmit motion vector
- Intra or Inter
  - Should MV found in step 1 be used? If so, select "Inter"
- Coded or Not coded
  - If quantized prediction error is zero, select "Not coded"
- Quant or No Quant
  - If quant scale needs to be changed, select "Quant"

35

MC / No_MC mode



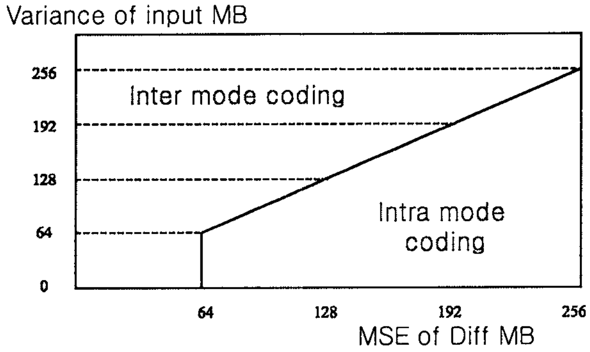
BD: Macroblock difference with no displacement

DBD: displaced block difference

36

Intra/inter decision

- Intra: DCT->Quant->VLC->transmit
- Inter: diff MB betwn current and previous frame MB is DCT coded
- Ex)



37

Intra/inter decision

- Variance of input MB

$$\mathrm{VAR} = \mathrm{E}\left[(x_0(m,n) - \bar{x}_0(i,j))^2\right]$$

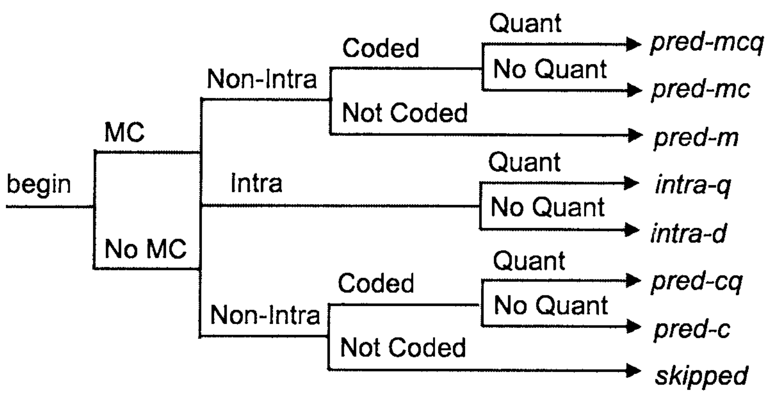
- MSE between input and MCed MB

$$\mathrm{MSE} = \mathrm{E}\left[(x_0(m,n) - x_{mc}(m,n))^2\right]$$

- In Inter mode, MC or No_MC mode can be used
- If MSE large, scene change or fast motion, and inter-picture prediction is ineffective, ->Intra coding
- If MSE too small, no change between consecutive pictures and do not need to transmit the MB

38

Example of MB Type selection for P-pictures



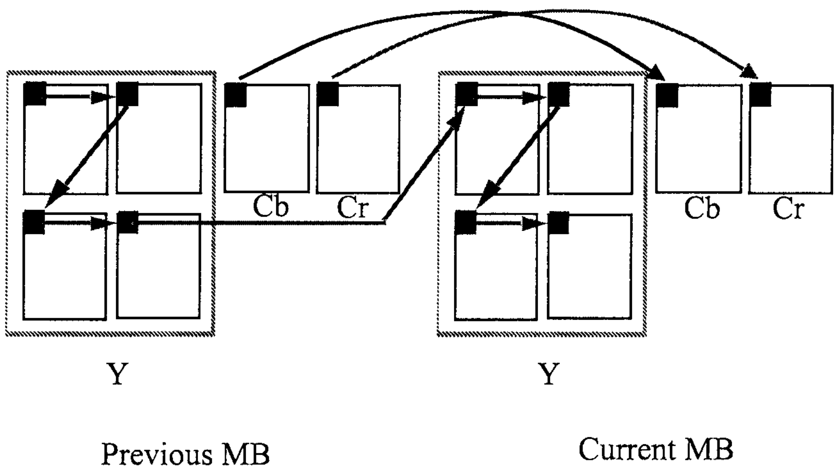
39

Entropy Coding

- Statistical redundancy reduction
- VLC/FLC
- DCT coef : VLC
  - INTRA coded DC : DPCM
  - INTRA coded & INTER coded AC :
    Huffman coding, [zero run, QC(i,j)]
  - Zig-zag scanning
- Motion Vector : VLC(Huffman code)
- MB Address, MB type, Coded block pattern use all VLC

40

DC Prediction



41

GOP and picture ordering

- The coded bitstream is transmitted in decoding order and the decoder converts it into display order
- GOP is used to assist random access
- The first coded picture in a GOP is an I picture
- The last coded picture in display order is either I or P
- Example) M=3, N=9, GOP_rate = frame_rate/N

Display order	I 0	B 1	B 2	P 3	B 4	B 5	P 6	B 7	B 8	I			
Decoding order	I 0	P 3	B 1	B 2	P 6	B 4	B 5	I	B 7	B 8			

42

CBR Rate control

- To achieve a constant bit rate, a buffer is used to smooth out high variability in bits/frame.
- In practice, I frames are often given highest quality, since they from the basis of prediction for all other pictures in the GOP.
- As complexity increases, the quant scale, on average, is increased to avoid buffer overflow.
- To approach constant quality <u>from frame to frame</u>, bits are "stolen" from simple frames and given to complex frames.
- To approach constant quality <u>within a frame</u>, bits are "stolen" from simple areas and given to complex areas.

43

Bit rate example

- Assume each picture is comprised of 10 slices at 30Hz frame rate. Number of bits per slice is given by
  1. M=2, N=15; I: 12,700 P: 5,600 B: 800
  2. M=3, N=15; I: 20,000 P: 6,400 B: 1,200
- Then total bit rate becomes 1.15 Mbps as:

1. I B P B P B P B P B P B P B P

(12,700 + 7 x 5600 + 7 x 800) x 10 x (30/15) =1.15Mbps

2. I B B P B B P B B P B B P B B

(20,000 + 4 x 6400 + 10 x 1200) x 10 x (30/15) = 1.15Mbps

44

MPEG-2

Differences from MPEG-1

- Search on fields, not just frames.
- 4:2:2 and 4:4:4 macroblocks
- Frame sizes as large as 16383 x 16383
- Scalable modes: Temporal, spatial, SNR,...
- Non-linear macroblock quantization factor
- A bunch of minor fixes

45

MPEG-1 and MPEG-2

	MPEG-1	MPEG-2
Video format	SIF progressive	SIF, 4:2:0, 4:2:2, 4:4:4, P/I
Picture quality	VHS	Distribution/contribution
Bit rate	Variable(< 1.856Mbps)	Variable(< 100Mbps)
Low delay mode	< 150ms	< 150ms(when no B)
Accessibility	Random access	Random access
Scalability	n/a	SNR, spatial, temporal
Compatibility	n/a	Forward, backward, up, down
Transmission error	Error protection	Error resilience
Editing bitstream	Yes	Yes
DCT	Noninterlaced	Field(prog) or frame(interlaced)
Motion estimation	Noninterlaced	Field, frame, dual-prime
Motion vectors	MV for P, B	For P, B, & concealment MV
Scanning	zigzag	Zigzag & alternate

46

MPEG-2

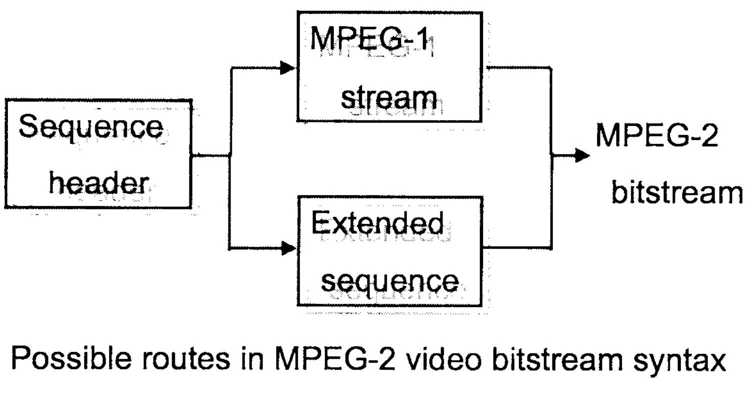
MPEG-2 target applications

Level	Size	Pixels/sec (Mbits)	bit-rate	Application
Low	352 x 240	3 M	4	consumer tape equiv.
Main	720 x 480	10 M	15	studio TV
High 1440	1440 x 1152	47 M	60	consumer HDTV
High	1920 x 1080	63 M	80	production

47

MPEG-2 video bitstream

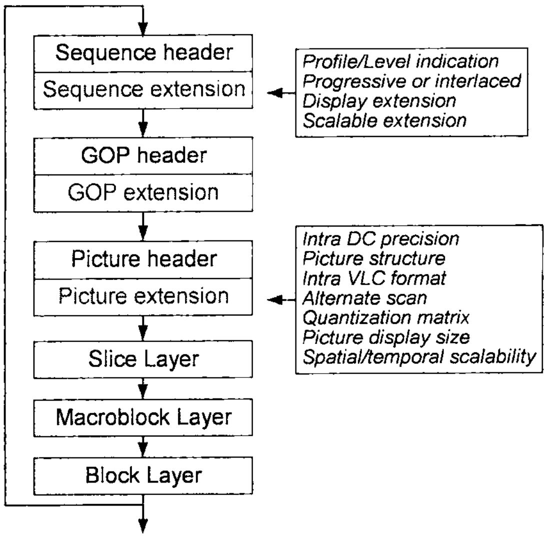
- Syntactic superset of MPEG-1 bitstream



Possible routes in MPEG-2 video bitstream syntax

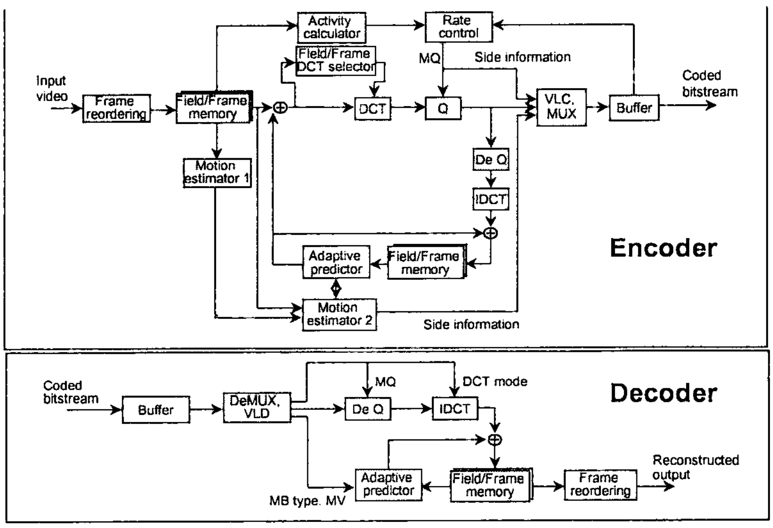
48

Extended MPEG-2 hierarchy



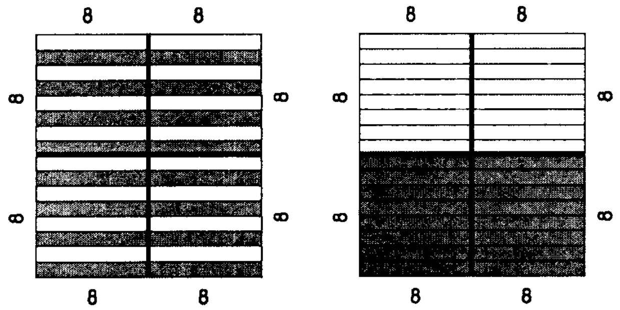
49

MPEG-2 codec



50

Frame /field block organization

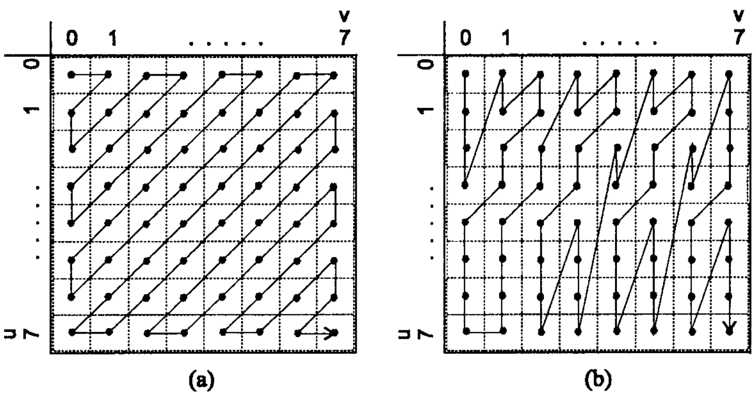


4 Frame blocks/MB
(two frames interlaced)

4 Field blocks/MB
(even, odd fields separately)

51

Alternate scan



Two scanning methods of the DCT coef in MPEG-2

(a) Zigzag scan, (b) alternate scan (due to less vertical correlation in field based prediction)

52

Prediction modes in MPEG-2

- Field prediction:
  - Top field has parity zero, bottom has parity one.
  - Prediction is made between the same parity fields or between the opposite parity fields
- Frame prediction
  - Even in this mode, field predictions can be selected on an MB-by-MB basis
- (16x8) motion compensation
  - 1^{st} MV applied to upper 16x8, 2^{nd} to lower 16x8
  - This is used with field pictures only
- Dual-prime prediction
  - MV refinement process in P pictures only
  - Predictions are made from 2 reference fields and averaged to form final prediction

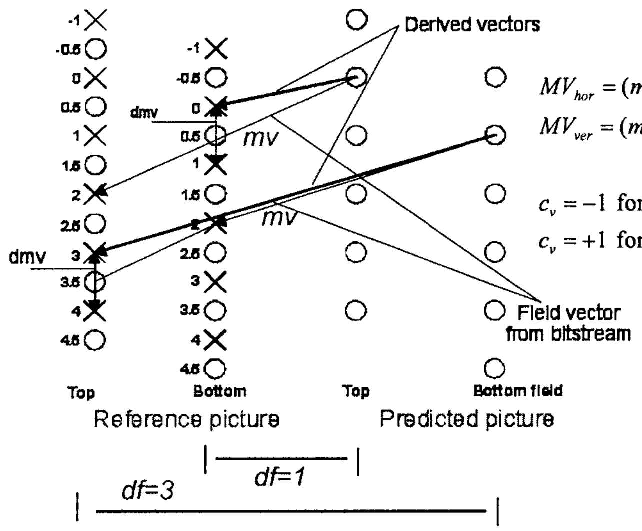
53

Prediction modes in MPEG-2

Field pictures	Frame pictures
Field-based	Frame-based
16 x 8 MC	Field-based
Dual-prime	Dual-prime

54

Dual-prime prediction



$$MV_{hor} = (mv_{hor} \times d_f)//2 + dmv_{hor}$$

$$MV_{ver} = (mv_{ver} \times d_f)//2 + c_v + dmv_{ver}$$

$c_v = -1$ for top-to-bottom prediction

$c_v = +1$ for bottom-to-top prediction

55

Profiles and Levels

- **Problem : A Decoder that could decode *any* MPEG-2 bitstream would be prohibitive in terms of memory and performance.**
  Decoder manufacturers might choose proprietary subsets of the syntax, preventing interoperability.
- **Solution : Pre-defined subsets of the syntax : Profiles & Levels create "compliance points"**
- ***Profile* : A defined subset of syntax elements in MPEG-2 (e.g, 4:2:0 only, I/P frames only, field DCT, etc.)**
- ***Level* : Parameter constraints on those syntax elements (e.g., max Picture Size, max Bit Rate, max Vertical Motion Vector, max Buffer Size, etc.)**

56

Profiles and Levels

- **Profiles : Simple, Main, SNR, Spatial, High, 4:2:2**
- **Levels : Low, Main, High-1440, High**

  ***Not all Profile/Level combinations are allowed.**
- **Main Profile :** - B Frames supported (not so in Simple Profile)

  - 4:2:2 and 4:4:4, and Scalable Modes not supported

  - Restricted slice structure
- **Main Level :** - max Picture size : 720 576, 30 frames/sec

  - max Bitrate : 15 Mbps

  - max Buffer size : 1.835008 Mbits
- **A Compliance Point is a Profile at a Level,**
  - e.g., Main Profile at Main Level, "MP@ML"

57

MPEG-2 scalability

- Scalability allows a subset of a bitstream to be decoded into meaningful imagery
- Also useful for error resilience on prioritized transmission media
- Some loss of efficiency happens due to extra overhead
- Scalable modes - Spatial

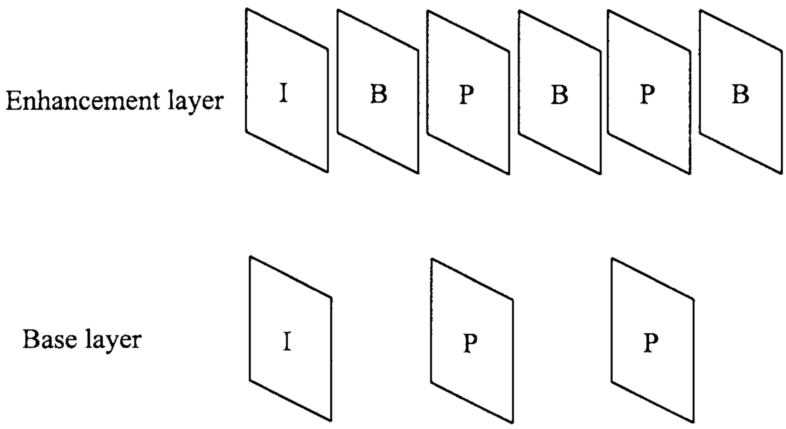
  - Temporal

  - SNR
- Certain combinations of these modes are also allowed, with maximum of two enhancement layers.

58

Temporal scalability

- Different temporal resolution



59

Spatial scalability

- Forward/Backward Compatibility

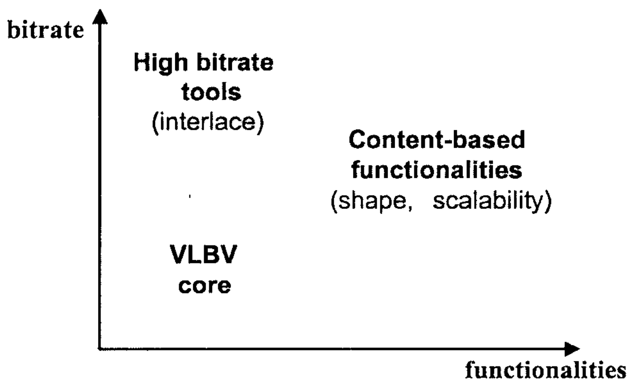
60

SNR scalability

61

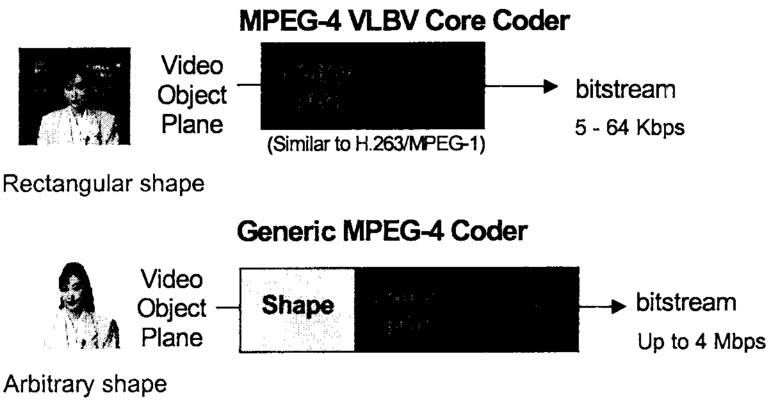
MPEG-4 algorithm & tools

- Classification of bitrates and functionalites



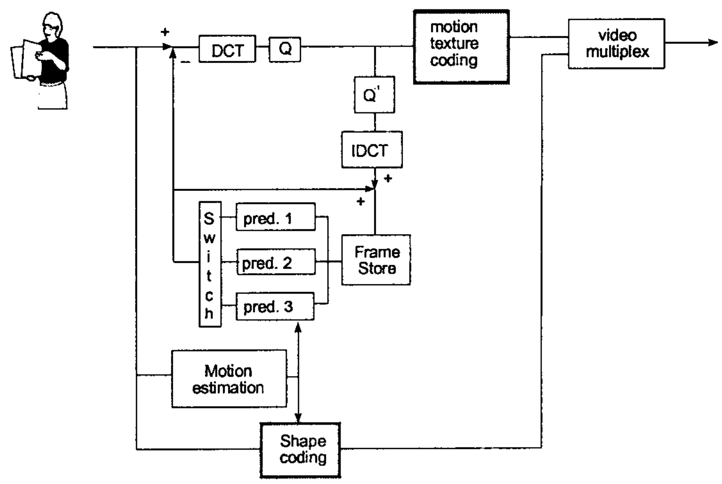
62

VLBV core and generic coder
(VLBV: Very Low Bitrate Video)



63

MPEG-4 video coder



64

Data structure in MPEG-4

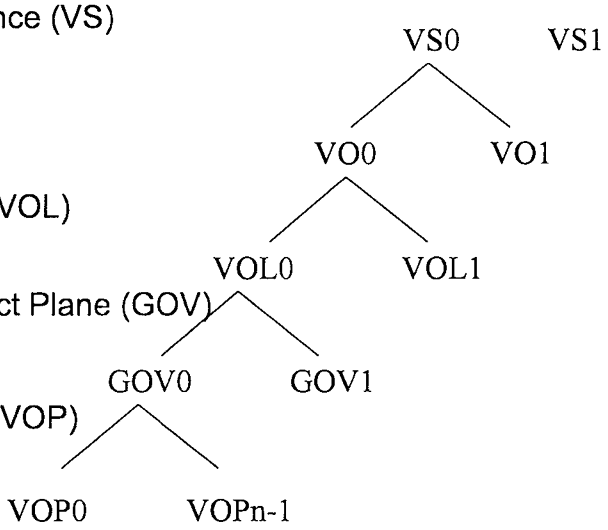
Visual object Sequence (VS)

Video Object (VO)

Video Object Layer (VOL)

Group of Video Object Plane (GOV)

Video Object Plane (VOP)



65

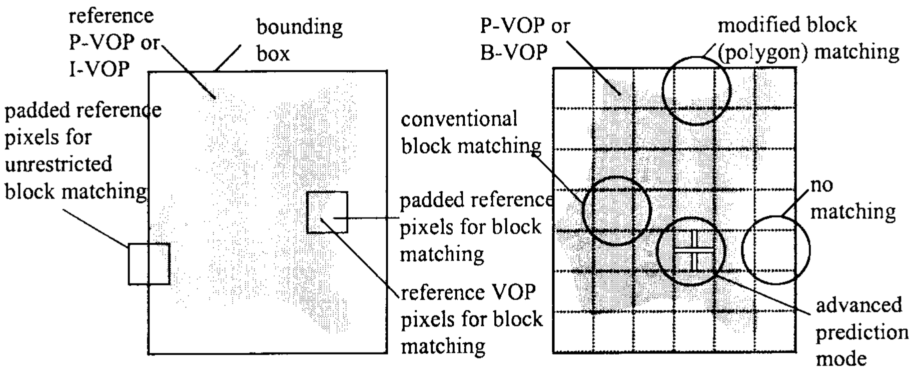
List of major natural video tools

- Binary shape
- Padding
- Motion compensation
- Quantization
- AC/DC prediction
- Scanning
- I, P, B modes
- Temporal scalability
- Spatial Scalability
- Error resilience
- Static sprites
- Interlaced coding
- 12-bit video
- Static texture

Overlapped motion compensation
Advanced motion compensation
Method 1
Method 2
Non-linear
Type 1
Type 2
Slice synchronization
Extended header code
Data partitioning
Reversible VLC
Basic
Low delay
Scalable

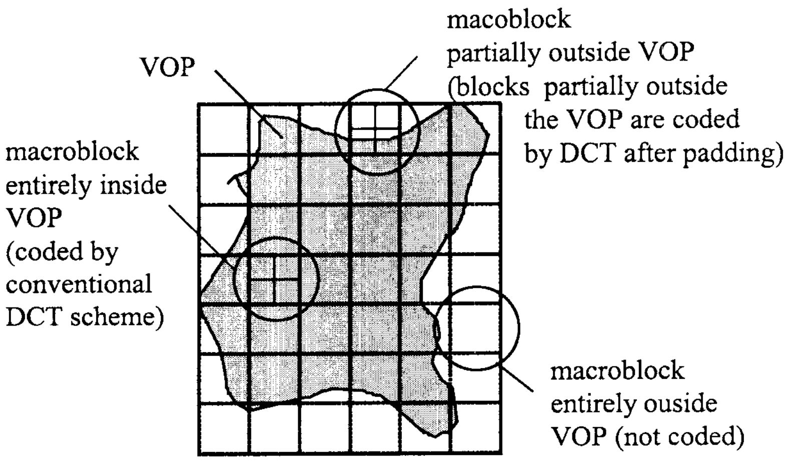
66

Motion compensation



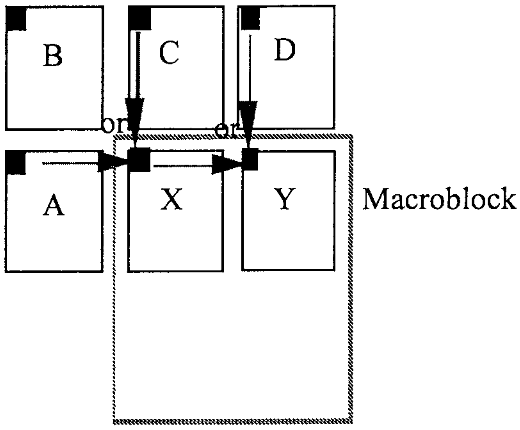
67

Texture coding tools



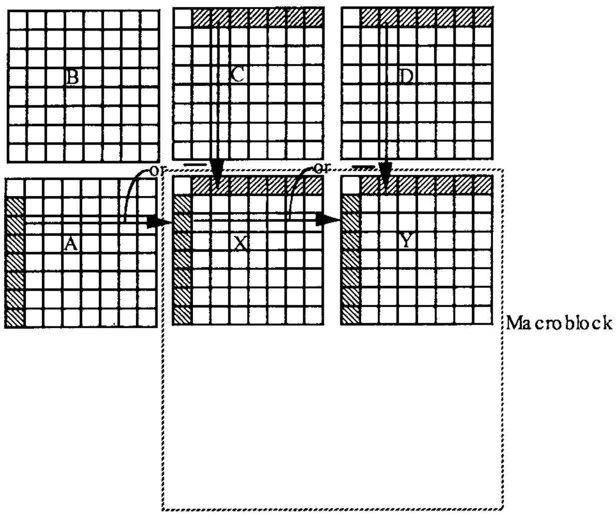
68

Adaptive DC prediction



69

Adaptive AC prediction



70